DOCUMENT RESUME

ED 064 355 TM 001 523

AUTHOR Hoffman, Richard J.
TITLE The Efficiency Index in Item Analysis.
PUB DATE Apr 72
NOTE 39p.; Paper presented at the annual meeting of the American Educational Research Association (Chicago, Illinois, April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Evaluation Methods; *Item Analysis; *Mathematical Concepts; *Probability; *Statistical Analysis; Test Interpretation

ABSTRACT

In this paper, a new item analysis index, e, is derived as a function of difficulty and discrimination to represent item efficiency. Item discrimination in this paper is not independent of item difficulty, and it is demonstrated algebraically that the maximum discriminating power of an item may be determined from its difficulty. Item efficiency is defined as the ratio of observed discrimination to maximum discrimination. The e-index will range from zero to unity and will provide additional information for item analyses. Its probability interpretations may provide an attractive psychometric criterion for the retention or rejection of items. (Author)

THE EFFICIENCY INDEX IN ITEM ANALYSIS

Richard J. Hofmann

Department of Educational Psychology

Miami University

Oxford, Ohio

Paper Presented to
Section D
American Educational Research Association
April 3-7, 1972

ABSTRACT

In this paper a new item analysis index, e, is derived as a function of difficulty and discrimination to represent item efficiency. In this paper item discrimination is not independent of item difficulty and it is demonstrated algebraically that the maximum discriminating power of an item may be determined from its difficulty. Item efficiency is defined as the ratio of observed discrimination to maximum discrimination.

The e-index will range from zero to unity and will provide additional information for item analyses. Its probability interpretations may provide an attractive psychometric criterion for the retention or rejection of items.

In a typical analysis of a test item two indices are usually computed, a difficulty index and a discrimination index. If one assumes an analysis based upon the performance of two groups on the item, typically referred to as a U-L analysis, then a two by two contingency table may be used in the tabulation of the indices. Such an approach would typify the approaches suggested by Kelley (1939), Johnson (1951) and Cureton (1957) and is discussed in almost any basic measurement text devoting some space to item analyses.

Assume that N' individuals have responded to some item in either a positive fashion, r, or a negative fashion, w. Furthermore assume that either on the basis of their total scores on the instrument associated with the item or on the basis of some outside criterion two equal groups, g_1 and g_2, are determined from N'. In this case g_1 and g_2 in total represent N individuals where N may be equal to or proportionate to N'. Then using the subscripts 1 and 2 to denote those symbols associated with g_1 and g_2, respectively, the N responses to the item are presented in Table 1.

Table 1

Contingency Table Summarization of Two Group Responses to a Single Item

Response Category	Group g_1	Group g_2	Total
Positive	r_1	r_2	r_1+r_2
Negative	w_1	w_2	w_1+w_2
Total	r_1+w_1	r_2+w_2	N

The difficulty index, a, of the item may be denoted as:

$$\underline{a} = \frac{\underline{r}_1+\underline{r}_2}{\underline{N}} \qquad [1]$$

and it may range from zero to unity. One interpretation of $\underline{a}$ is that it represents the proportion of $\underline{N}$ individuals responding positively to the item. A second interpretation is that $\underline{a}$ represents the probability of observing a positive response to the item, $\underline{P}(\underline{r})$.

The discrimination index of the item, $\underline{b}$, may be denoted as:

$$\underline{b} = \frac{\underline{r}_1}{\underline{r}_1+\underline{w}_1} - \frac{\underline{r}_2}{\underline{r}_2+\underline{w}_2}.$$

Where

$$\underline{r}_1+\underline{w}_1 = \underline{r}_2+\underline{w}_2 = \underline{n};$$

thus

$$\underline{b} = \frac{\underline{r}_1}{\underline{n}} - \frac{\underline{r}_2}{\underline{n}} \qquad [2]$$

and it may range from positive to negative unity. One interpretation of $\underline{b}$ is that it represents the difference between two conditional probabilities, the probability of a correct response given membership in group one, $\underline{P}(\underline{r}|\underline{g}_1)$, less the probability of a correct response given membership in group two, $\underline{P}(\underline{r}|\underline{g}_2)$.

In a very special sense the marginals of Table 1 are fixed and there is an interdependence between difficulty and discrimination. In this paper difficulty is chosen to be an independent variable while discrimination is chosen to be the dependent variable. That is to say, discrimination is assumed to be a function of difficulty and thus its magnitued for any item is tempered by the magnitude of the item's difficulty index.

A frequent problem encountered by "users" of item analyses is one of interpreting both indices, difficulty and discrimination, simultaneously and making a decision about the disposition of an item, either retaining or rejecting it for future use. All too frequently interpretations are confused when either or both indices depart, even slightly, from .50 for difficulty and positive or negative unity for discrimination. (It should be noted here that it is a popular misconception that the ideal difficulty index should be .50. For a comprehensive discussion of this point see Henrysson, 1971.)

The major objective of this paper is one of deriving a new index that will facilitate the interpretation of item analyses. In this paper a new index, $\underline{e}$, is presented as a function of difficulty and discrimination. Conceivably the $\underline{e}$ index might facilitate as much if not more information than the simultaneous interpretations of difficulty and discrimination while at the same time it should be less confusing as it may be interpreted within a simple probability framework.

Maximum Discrimination as a Function of Difficulty

As noted the discrimination index may be thought of as being dependent upon the difficulty index. When discussed generically the discrimination index is incorrectly assumed to range from positive to negative unity. Regardless of the type of discrimination index used the absolute maximum value of unity can be attained only when the associated difficulty index is .50.

As the difficulty index of an item deviates from .50, either above or below it, the maximum ceiling of the discrimination index is reduced from unity. For each metric unit of deviation from .50 for a difficulty index there is a two metric unit reduction from unity for the maximum ceiling of the associated discrimination index. (See Hofmann, 72, for greater detail on this.) Thus given a difficulty index of $\underline{a}$ its absolute deviation from .50, $|\underline{D}|$, may be used to compute the ceiling or maximum possible discrimination index, in absolute value terms $|\underline{b}^*|$, of the associated discrimination index.

$$|\underline{D}| = |.50 - \underline{a}| \tag{3}$$

and

$$|\underline{b}^*| = 1.00 - 2|\underline{D}|. \tag{4}$$

Logically the principle involved in computing maximum discrimination is presented by equations 3 and 4, however the pragmatics of the concept are obscured by the equations. A more reasonable set of equations for maximum discrimination may be utilized if one is willing to be cognizant of the direction of the difficulty deviation. Assume that ($\underline{a} \leq .50$) then maximum discrimination may be defined algebraically as:

$$\underline{b}^* = 2\underline{a}. \tag{5}$$

Assume that ($\underline{a} \geq .50$) then maximum discrimination may be defined as:

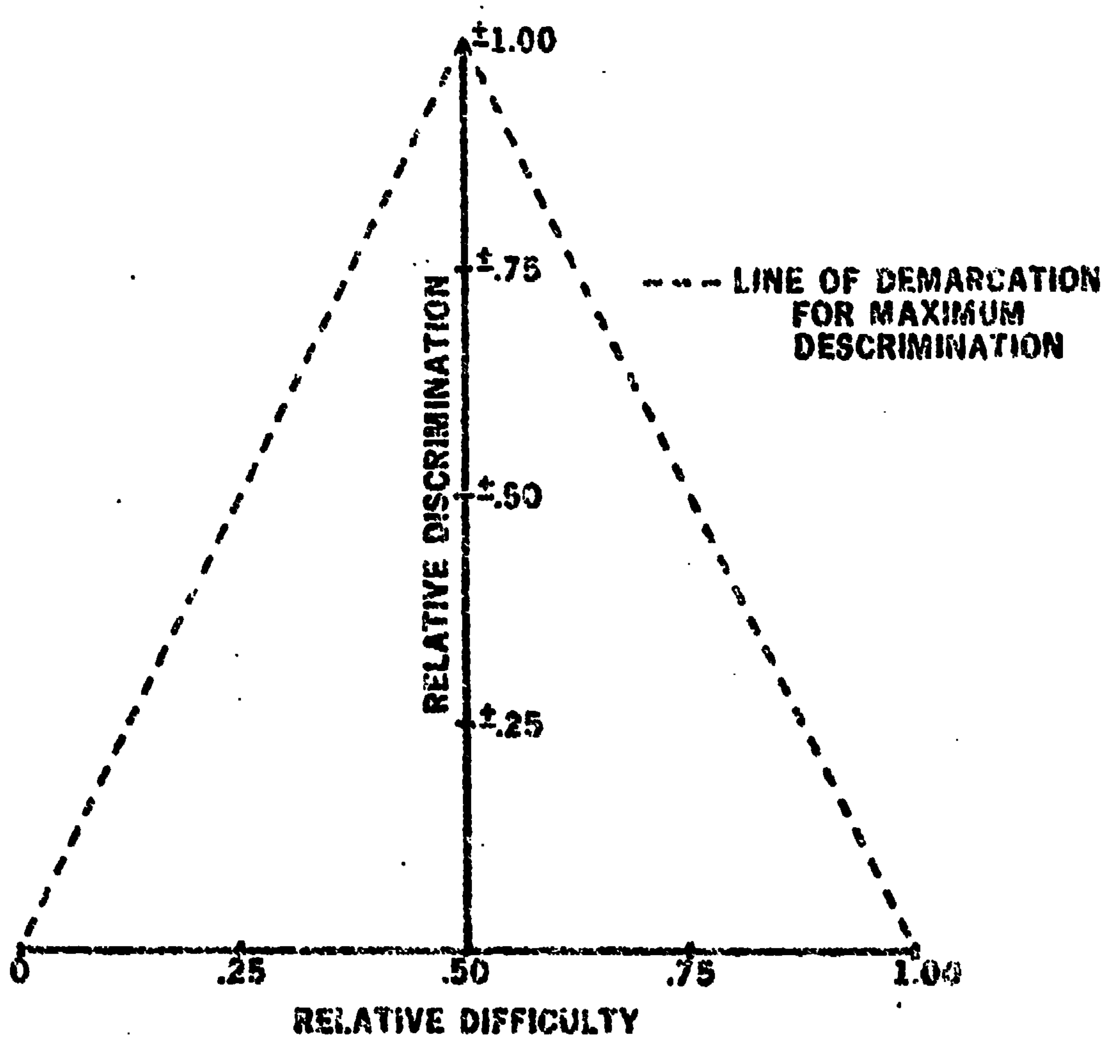
$$\underline{b}^* = 2(1-\underline{a}). \tag{6}$$

Geometrically maximum discrimination has a perfect curvilinear relationship to difficulty within the four quadrants of a two dimensional space. Within any one quadrant maximum discrimination is linearly related to difficulty. Because of an isomorphism between quadrants the nonadjacent quadrants are reflections of each other and any pair of adjacent quadrants may be used to depict the relationship between difficulty and maximum discrimination.

In Figure 1 two adjacent quadrants of a Cartesian coordinate system have been depicted. The abcissa of this system represents a difficulty continuum while the ordinate of the system represents a "dual signed" discrimination continuum, the values may be interpreted as either positive or negative.

The origin is denoted on the difficulty continuum as a .50 so that any movement along the continuum will represent directed deviations from .50. The dashed line of demarcation within the "left quadrant" represents the line that is defined by any set of coordinates ($\underline{a}$,$\underline{b}^*$) where ($\underline{a} \leq .50$) and $\underline{b}^*$ is a maximum discrimination value defined on the discrimination continuum. The dashed line in the "right quadrant" is the line that is defined by any set of coordinates ($\underline{a}$,$\underline{b}$*) where ($\underline{a} \geq .50$) and $\underline{b}$* is a maximum discrimination index, a value defined on the discrimination continuum.

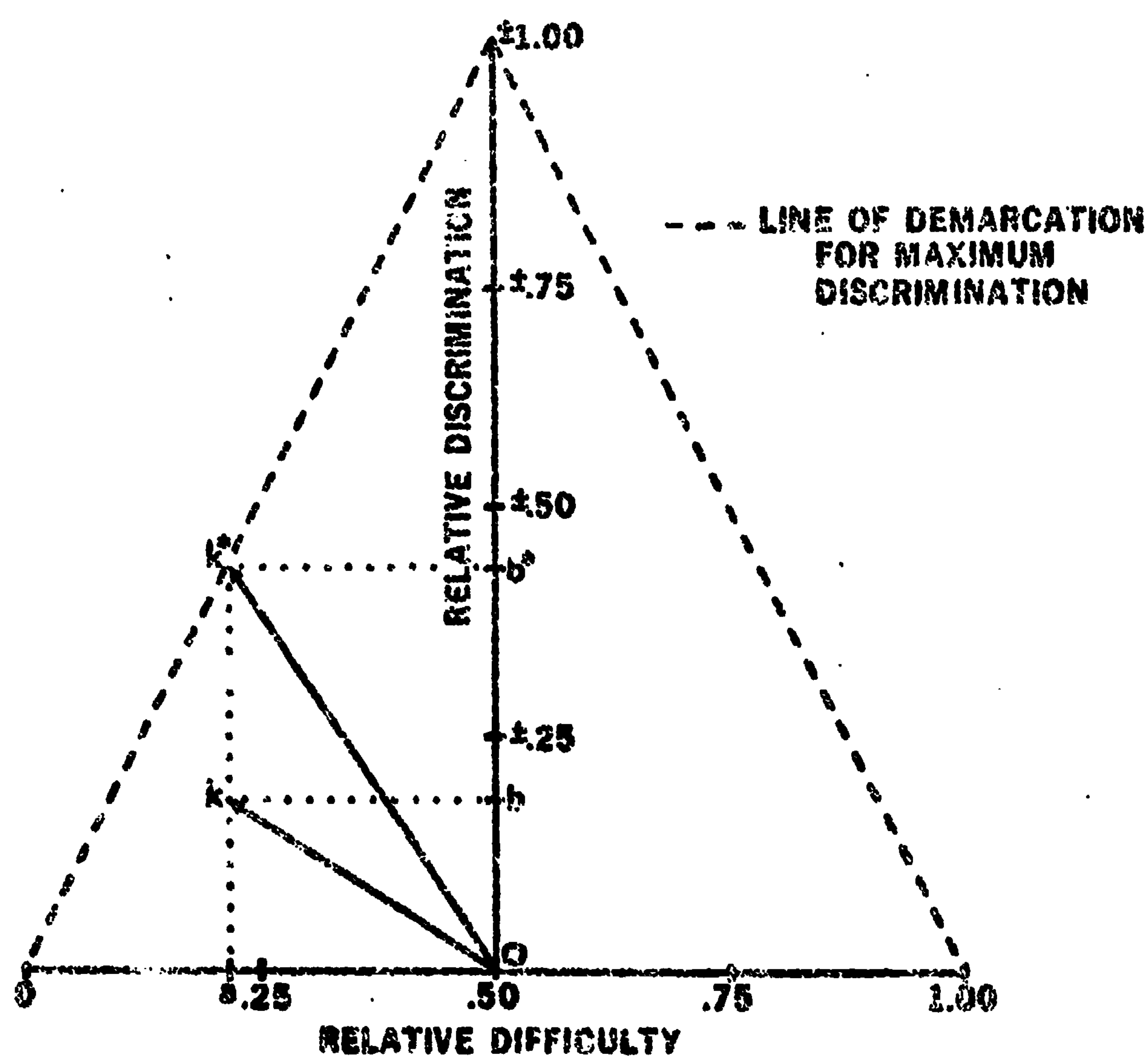
Figure 1. General Cartesian Co-ordinate System
Defined by Difficulty and Discrimination



In Figure 2 the terminus of an item vector, k, has been plotted with respect to the difficulty of the item, a, and the discrimination of the item, b. A second item vector, k*, has been plotted with respect to the difficulty of the item, a, and the maximum discrimination, b*, of the item.

Assume the Roman letter o is associated with the origin. There are two right triangles depicted in Figure 2, ako and ak*o. Both triangles have the same base, the deviation of the difficulty index a from .50. The triangle defined by ak*o, being the ideal triangle, will always be larger than triangle ako, the observed triangle. The ratio of the areas of the two triangles will indicate the size of ako relative to the maximum size it might have obtained. That is, the ratio of the two areas may be thought of as representing the efficiency of the item. The better an item functions the more closely will the ratio of the two areas approach unity.

Figure 2. Item Vector k Defined by Co-ordinates (a,b) and Ideal Item Vector k* Defined by Co-ordinates (a,b*).



Technically the area of the observed triangle is a function of the types of discriminations the item makes. In referring to Table 1 it may be observed that in actuality (r_1+r_2) individuals are judged as being better than (w_1+w_2) other individuals. However, the item functions as though $[(r_1+r_2)(w_1+w_2)]$ dichotomous discriminations are made. The absolute maximum number of discriminations which may be made for a given item is, $(N /4)$.

Consider the component parts of the equation

$$(r_1+r_2)(w_1+w_2) = r_1w_1 + r_1w_2 + r_2w_1 + r_2w_2; \qquad [7]$$

then it is possible to consider the concept of "proper" and "improper" discriminations. Let the term proper discriminations refer to those point discriminations which are desirable in the sense that they result through a maximizing of the frequency of one particular response type, positive response, in one particular group, group one, while the other response type, negative response, is being maximized in the other group, group two. The term improper discrimination may be associated with those point discriminations which are not desirable in the sense that they occur as a result of undesirable response types occurring in both groups. For any item the number of proper discriminations is characterized by (r_1w_2) and the improper discriminations by (r_2w_1). Implicity for easy items the negative responses should all be accrued by group two, w_2, and for difficult items the positive responses should all be accrued by group one, r_1.

A discrimination index in terms of point discriminations, $\{b\}$, is just the difference between proper and improper discriminations.

$$\{b\} = (r_1w_2) - (r_2w_1) \qquad [8]$$

and the relative discrimination index is given by

$$\underline{b} = \frac{\{\underline{b}\}}{\frac{\underline{N}^2}{4}} \qquad [9]$$

The maximum discrimination index, however, assumes no improper discriminations. Either $\underline{r}_2$ or $\underline{w}_1$ is assumed to be zero and either $\underline{r}_1$ or $\underline{w}_2$ is assumed to be $\underline{n}$. Thus, maximum discrimination in terms of point discriminations, $\{\underline{b}^*\}$, represents the maximum number of proper discriminations possible for a given $\underline{N}$ and difficulty index.

$$\{\underline{b}^*\} = (\underline{r}_1+\underline{r}_2)\underline{n};\ \underline{a} \leq .50 \qquad [10\text{-a}]$$

or

$$\{\underline{b}^*\} = (\underline{w}_1+\underline{w}_2)\underline{n};\ \underline{a} \geq .50 \qquad [10\text{-b}]$$

and the relative maximum discrimination index is given by

$$\underline{b}^* = \frac{\{\underline{b}^*\}}{\frac{\underline{N}^2}{4}} \qquad [11]$$

Now given that the area of any triangle is equal to one-half the base multiplied by the altitude and given that both triangles in Figure 2 have the same base then the ratio of their areas is algebraically equivalent to the ratio of their altitudes. Alternatively the ratio represents the number of observed proper discriminations less observed improper discriminations divided by the maximum possible number of proper discriminations. The ratio of $\{\underline{b}\}$ to $\{\underline{b}^*\}$ will range from zero to unity, assuming $\{\underline{b}\}$ is positive, and may be thought of, conceptually, as representing the "purity" of the discriminations made or the efficiency of the item. Let $\underline{e}$ represent a general efficiency index then:

$$\underline{e} = \frac{\{\underline{b}\}}{\{\underline{b}^*\}} \qquad [12]$$

and in modified form:

$$\underline{e} = \frac{\underline{b}}{\underline{b}^*} \qquad [13]$$

When one considers the fact that ($\underline{b}^*$) will always be positive and ($\underline{b}$) may be positive or negative then it becomes immediately apparent that $\underline{e}$ as defined by equations 12 and 13 may be positive or negative. When $\underline{e}$ is negative it is negative because more improper than proper discriminations were made. The terms proper and improper were somewhat arbitrarily assigned to two quantities on the assumption that more positive responses and, hence, fewer negative responses would always be made by group one relative to group two. When this assumption is not met the $\underline{e}$ index will be negative. However, for interpretations within the framework of proportions and areas the sign of $\underline{e}$ may be neglected. The negative sign of $\underline{e}$ becomes meaningful only within the framework of probability.

Given the conditional magnitude of the difficulty index the general $\underline{e}$ may be further specified as:

$$\underline{e}_1 = \frac{\underline{b}}{2\underline{a}} \;;\; \underline{a} \leq .50 \qquad [14]$$

general efficiency for items having difficulty indices less than or equal to .50 and:

$$\underline{e}_2 = \frac{\underline{b}}{2(1-\underline{a})} \;;\; \underline{a} \geq .50 \qquad [15]$$

general efficiency for items having difficulty indices greater than or equal to .50.

Certain initial observations may be made with respect to $\underline{e}$ and proportion interpretations

(a) If the observed discrimination index of an item is zero then the efficiency of the item is zero.

(b) For any level of difficulty, excluding zero and unity, it is theoretically possible for $\underline{e}$ to range from zero to unity assuming a positive discrimination index.

(c) Efficiency is the ratio of observed proper discriminations less improper discriminations to the maximum possible number of proper discriminations for a given difficulty level and group size.

(d) The general index $\underline{e}$ is indicative of how well an item has functioned relative to how well it might have functioned for a given $\underline{N}$ and specific difficulty level.

Probability Interpretations of Efficiency

In the previous section it was noted that the general efficiency index could be subdivided into two indices, one for items having difficulties less than or equal to .50, henceforth efficiency of the first kind, $\underline{e}_1$, and one for items having difficulties greater than or equal to .50, henceforth efficiency of the second kind, $\underline{e}_2$. The indices of efficiency may be further utilized to make probability interpretations with respect to positive responses and with respect to negative responses.

Equation 14 defining $\underline{e}_1$, for items having difficulties less than or equal to .50, may be modified to define a computational equation for $\underline{e}_1$ regardless of item difficulty and sign of the discrimination index.

$$\underline{e}_1 = \frac{\underline{r}_1 - \underline{r}_2}{\underline{r}_1 + \underline{r}_2} \qquad [16]$$

Similarly equation 15 defining $\underline{e}_2$, for items having difficulties greater than or equal to .50, may be modified to define a computational equation for $\underline{e}_2$ regardless of item difficulty and sign of the discrimination index.

$$\underline{e}_2 = \frac{\underline{w}_2-\underline{w}_1}{\underline{w}_1+\underline{w}_2} \qquad [17]$$

Utilizing equations 16 and 17 it is possible to discuss conditional probabilities and note that quite unlike traditional U-L discrimination indices efficiency considers two events which are mutually exclusive and exhaustive with respect to a given sample space.

Assume that a positive response has been made to an item. Given an individual making a positive response the probability that the individual is a member of $\underline{g}_1$ is given by $\underline{P}(\underline{g}_1|\underline{r})$ while the probability that the individual is a member of $\underline{g}_2$ is given by $\underline{P}(\underline{g}_2|\underline{r})$, where:

$$\underline{P}(\underline{g}_1|\underline{r}) = \frac{\underline{r}_1}{\underline{r}_1+\underline{r}_2} \qquad [18]$$

and

$$\underline{P}(\underline{g}_2|\underline{r}) = \frac{\underline{r}_2}{\underline{r}_1+\underline{r}_2}. \qquad [19]$$

Then

$$\underline{e}_1 = \underline{P}(\underline{g}_1|\underline{r}) - \underline{P}(\underline{g}_2|\underline{r}) \qquad [20]$$

efficiency of the first kind is the difference between two conditional probabilities, where the probabilities are for group membership, either $\underline{g}_1$ or $\underline{g}_2$, given a positive response.

Assume that a negative response has been made to an item. Given an individual making a negative response to an item the probability that the individual is a member of $\underline{g}_1$ is given by $\underline{P}(\underline{g}_1|\underline{w})$ while the probability that the individual is a member of $\underline{g}_2$ is given by $\underline{P}(\underline{g}_2|\underline{w})$, where:

$$\underline{P}(\underline{g}_1|\underline{w}) = \frac{\underline{w}_1}{\underline{w}_1+\underline{w}_2}. \qquad [21]$$

and

$$\underline{P}(\underline{g}_2|\underline{w}) = \frac{\underline{W}_2}{\underline{W}_1+\underline{W}_2} \qquad [22]$$

Then:

$$\underline{e}_2 = \underline{P}(\underline{g}_2|\underline{w}) - \underline{P}(\underline{g}_1|\underline{w}) \qquad [23]$$

efficiency of the second kind is the difference between two conditional probabilities, the probability of group membership given a negative response.

Efficiency of the first kind and efficiency of the second kind are both mutually exclusive and exhaustive with respect to sample space:

$$\underline{e}_1 = \underline{P}(\underline{g}_1|\underline{r}) - \underline{P}(\underline{g}_2|\underline{r}); \qquad [24]$$

$$1.0 = \underline{P}(\underline{g}_1|\underline{r}) + \underline{P}(\underline{g}_2|\underline{r}); \qquad [25]$$

$$\underline{e}_2 = \underline{P}(\underline{g}_2|\underline{w}) - \underline{P}(\underline{g}_1|\underline{w}); \qquad [26]$$

$$1.0 = \underline{P}(\underline{g}_2|\underline{w}) + \underline{P}(\underline{g}_1|\underline{w}). \qquad [27]$$

Following logically from these equations are:

$$\underline{P}(\underline{g}_1|\underline{r}) = \frac{\underline{e}_1+1}{2}; \qquad [28]$$

$$\underline{P}(\underline{g}_2|\underline{r}) = \frac{1-\underline{e}_1}{2}; \qquad [29]$$

$$\underline{P}(\underline{g}_1|\underline{w}) = \frac{1-\underline{e}_2}{2}; \qquad [30]$$

$$\underline{P}(\underline{g}_2|\underline{w}) = \frac{\underline{e}_2+1}{2}. \qquad [31]$$

Such additional probability interpretations for efficiency should be quite appealing for certain in depth analyses of items. In Table 2 values for equations 28-31 are given for 100 efficiency indices between zero and unity.

For a negative $\underline{e}$ column headings may be interchanged so that the first column becomes the probability of the second event and the second column becomes the probability of the first even. Similarly the odds columns should be interchanged.

Table 2

Probability Interpretations Associated with 100
Efficiency Indices between Zero and Unity

e	Probability of First Event	Probability of Second Event	Odds for First Event	Odds against First Event
.00	.500	.500	1.000	1.000
.01	.505	.495	1.020	.980
.02	.510	.490	1.040	.960
.03	.515	.485	1.061	.941
.04	.520	.480	1.083	.923
.05	.525	.475	1.105	.904
.06	.530	.470	1.127	.886
.07	.535	.465	1.150	.869
.08	.540	.460	[illegible]	.851
.09	.545	.455	1.197	.834
.10	.550	.450	1.222	.818
.11	.555	.445	1.247	.801
.12	.560	.440	1.272	.785
.13	.565	.435	1.298	.769
.14	.570	.430	1.325	.754
.15	.575	.425	1.352	.739
.16	.580	.420	1.280	.724
.17	.585	.415	1.409	.709
.18	.590	.410	1.439	.694
.19	.595	.405	1.469	.680
.20	.600	.400	1.500	.666
.21	.605	.395	1.531	.652
.22	.610	.390	1.564	.634
.23	.615	.385	1.597	.626
.24	.620	.380	1.631	.612
.25	.625	.375	1.666	.600
.26	.630	.370	1.702	.587
.27	.635	.365	1.712	.574

Table 2 (cont.)

e	Probability of First Event	Probability of Second Event	Odds for First Event	Odds against First Event
.28	.640	.360	1.777	.562
.29	.645	.355	1.816	.550
.30	.650	.350	1.857	.538
.31	.655	.345	1.898	.526
.32	.660	.340	1.941	.515
.33	.665	.335	1.985	.503
.34	.670	.330	2.030	.492
.35	.675	.325	2.076	.481
.36	.680	.320	2.125	.470
.37	.685	.315	2.124	.459
.38	.690	.310	2.225	.449
.39	.695	.305	2.278	.438
.40	.700	.300	2.333	.428
.41	.705	.295	2.389	.418
.42	.710	.290	2.448	.408
.43	.715	.285	2.508	.401
.44	.720	.280	2.571	.398
.45	.725	.275	2.636	.388
.46	.730	.270	2.703	.369
.47	.735	.265	2.773	.360
.48	.740	.260	2.846	.351
.49	.745	.255	2.921	.342
.50	.750	.250	3.000	.333
.51	.755	.245	3.081	.324
.52	.760	.240	3.166	.315
.53	.765	.235	3.255	.307
.54	.770	.230	3.347	.298
.55	.775	.225	3.444	.290
.56	.780	.220	3.545	.282
.57	.785	.215	3.651	.273

Table 2 (cont.)

e	Probability of First Event	Probability of Second Event	Odds for First Event	Odds against First Event
.58	.790	.210	3.761	.265
.59	.795	.205	3.878	.257
.60	.800	.200	4.000	.250
.61	.805	.195	4.128	.242
.62	.810	.190	4.263	.234
.63	.815	.185	4.405	.226
.64	.820	.180	4.555	.219
.65	.825	.175	4.714	.212
.66	.830	.170	4.882	.204
.67	.835	.165	5.060	.197
.68	.840	.160	5.250	.190
.69	.845	.155	5.451	.183
.70	.850	.150	5.666	.176
.71	.855	.145	5.896	.169
.72	.860	.140	6.142	.162
.73	.865	.135	6.407	.156
.74	.870	.130	6.692	.149
.75	.875	.125	7.000	.142
.76	.880	.120	7.333	.136
.77	.885	.115	7.695	.129
.78	.890	.110	8.990	.123
.79	.895	.105	8.523	.117
.80	.900	.100	9.000	.111
.81	.905	.095	9.526	.104
.82	.910	.090	10.111	.098
.83	.915	.085	10.764	.092
.84	.920	.080	11.500	.086
.85	.925	.075	12.333	.081
.86	.930	.070	13.285	.075
.87	.935	.065	14.384	.069

Table 2 (cont.)

e	Probability of First Event	Probability of Second Event	Odds for First Event	Odds against First Event
.88	.940	.060	15.666	.063
.89	.945	.055	17.181	.058
.90	.950	.050	19.000	.052
.91	.955	.045	21.222	.047
.92	.960	.040	24.000	.041
.93	.965	.035	27.571	.036
.94	.970	.030	32.333	.030
.95	.975	.025	39.000	.025
.96	.980	.050	49.000	.020
.97	.985	.015	65.666	.015
.98	.990	.010	99.000	.010
.99	.995	.005	199.000	.005
1.00	1.000	.000	---	---

It may be noted that e_1 and e_2 are proportional to each other. The ratio of e_2 to e_1 represents the odds in favor of a negative response while the ratio of e_1 to e_2 represents the odds in favor of a positive response.

$$\frac{e_1}{e_2} = \frac{w_1+w_2}{r_1+r_2} \qquad [32\text{-a}]$$

and

$$\frac{e_2}{e_1} = \frac{r_1+r_2}{w_1+w_2}. \qquad [32\text{-b}]$$

Such equations are referred to as "odd's ratios." The odds for and against a response type given a particular index are also reported for the 100 different e indices in Table 2.

Equations algebraically equivalent to equations 32-a and 32-b may be derived using the difficulty index.

$$\frac{e_2}{e_1} = \frac{a}{1-a} \qquad [33\text{-}a]$$

and

$$\frac{e_1}{e_2} = \frac{1-a}{a}. \qquad [33\text{-}b]$$

If one wishes to convert e_1 to e_2 it is only necessary to multiply e_1 by the odds in favor of a positive response, equation 33-a. The converse is true for converting e_2 to e_1.

The Probability of Obtaining an Observed e by Chance

The general e index has been discussed within the framework of e_1 and e_2. It was noted that for any given level of difficulty e may range from zero to unity. Quite logically one would like to know the probability of obtaining an observed e for any given index of difficulty. Generically, what is a significant e?

The model contingency table from which e is computed is unique within the framework of statistics. Theoretically e is a measure of departure from independence in the contingency table. However there is a different probability distribution for e associated with each uniquely different sample size, N, and each uniquely different difficulty level, a. In order to compute the probabilities associated with any given e it must be assumed that all four marginals of the contingency table are fixed. That is to say, the probabilities reported for any e are determined from the specific probability distribution of e associated with a given N and a.

Technically in order to test the null hypothesis of independence, (H_0:$e = 0$), it is necessary to compute the probability of obtaining the observed e and all possible e indices of a larger magnitude assuming constant marginals. Although Pearson's (1932) Chi square test might be used with this model of a contingency table it was not designed specifically for such use and in using it one would have to constantly keep in mind the consequences of its use with small sample sizes and also meet the assumptions of expected frequencies greater than five in the cells of the table.

A test designed specifically for the type of model contingency table associated with *e* is Fisher's (1935) exact test. Essentially Fisher's test would indicate the exact probability of obtaining an observed *e* given a particular *N* and *a*. Furthermore it could be used to compute the exact probabilities of each associated *e* greater than the observed *e*. In summing up all of these exact probabilities one would have the probability of obtaining an *e* as large as or larger than the one observed with the given level of difficulty, *a*, and group size, *N*.

For any test having more than four or five items or fifteen or sixteen individuals such an approach would be computationally time consuming.

However, it is possible to use a variation of Fisher's test, which is based upon the hypergeometric distribution, to establish the magnitudes of the *e* indices which would represent the extreme tails of the distributions of such indices for various difficulty levels and group sizes. Thus, it is possible, so to speak, to establish "tables of significance" for the *e* index by computing the probabilities associated with the most extreme values for *e* and work, computationally, toward the less extreme values.

The method suggested here is an approximation method whose numerical methods are reported by Hofmann (72). Essentially this variation involves the computation of the probability of the least independent contingency table, *e* = 1.0, for a given difficulty level and then continues to compute the probabilities for successively more independent tables, while the difficulty index is held constant. Such an approach avoids the extremely time consuming computation associated with large factorials, which will occur in the more dense, middle, area of a symmetric distribution. Further discussion of this procedure is beyond the scope of this paper. Using this procedure levels of significance for _e_ were computed for difficulty indices between zero and unity for samples ranging in size from 8 to 100 for the .05 level of significance and are reported in Table 3.

The interpretations associated with Table 3 are the same as those made for any percentile point, a one-tail test. For a given group, _N_, with a given difficulty, _a_, there is a symmetric distribution of possible _e_ indices. The tabled _e_ index is that _e_ associated with the percentile point which best approximates the .05 percentile point.

Table 3. Significant ($p \leq .05$) Values of e for Given Difficulty Levels and Group Sizes.

Difficulty	Total Group Size															
	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38
.99 - .01	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.98 - .02	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.97 - .03	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.96 - .04	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.94 - .05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.94 - .06	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.93 - .07	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.92 - .08	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.91 - .09	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.90 - .10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.89 - .11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.88 - .12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.87 - .13	-	-	-	-	-	-	-	-	-	-	-	1.00	-	-	-	1.00
.86 - .14	-	-	-	-	-	-	-	-	-	-	1.00	-	-	-	1.00	-
.85 - .15	-	-	-	-	-	-	-	-	-	1.00	-	-	-	1.00	-	-
.84 - .16	-	-	-	-	-	-	-	-	-	-	-	-	1.00	-	-	1.00
.83 - .17	-	-	-	-	-	-	-	-	1.00	-	-	1.00	-	1.00	1.00	-
.82 - .18	-	-	-	-	-	-	-	1.00	-	-	1.00	-	1.00	-	-	.71
.81 - .19	-	-	-	-	-	-	-	-	-	1.00	-	-	-	-	.71	-
.80 - .20	-	-	-	-	-	-	1.00	-	1.00	-	-	1.00	-	-	-	-
.79 - .21	-	-	-	-	-	-	-	-	-	-	1.00	-	-	.71	-	.75

Table 3 (cont.)

Difficulty	Total Group Size 8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38
.78 - .22	-	-	-	-	-	1.00	-	-	-	-	-	-	.71	-	.75	-
.77 - .23	-	-	-	-	-	-	-	1.00	-	1.00	-	.71	-	-	-	-
.76 - .24	-	-	-	-	-	-	-	-	-	-	-	-	-	.75	-	.78
.75 - .25	-	-	-	-	1.00	-	1.00	-	1.00	-	.71	-	.75	-	.78	-
.74 - .26	-	-	-	-	-	-	-	-	-	-	-	.75	-	-	-	.60
.73 - .27	-	-	-	-	-	-	-	1.00	-	.71	-	-	-	.78	-	-
.72 - .28	-	-	-	-	-	1.00	-	-	-	-	-	-	.78	-	.60	-
.71 - .29	-	-	-	1.00	-	-	-	-	.71	-	.75	-	-	.60	-	.64
.70 - .30	-	-	-	-	-	-	1.00	-	-	-	-	.78	-	-	-	-
.69 - .31	-	-	-	-	1.00	-	-	-	-	.75	-	-	.60	-	.64	-
.68 - .32	-	-	-	-	-	-	-	.71	-	-	.78	-	-	.64	-	.50
.67 - .33	-	-	1.00	-	-	1.00	-	-	.75	-	-	.60	-	-	.50	-
.66 - .34	-	-	-	-	-	-	-	-	-	-	-	-	.64	-	-	.54
.65 - .35	-	-	-	-	-	-	.75	-	-	.56	-	-	-	.50	-	-
.64 - .36	-	-	-	1.00	-	-	-	.75	-	-	.60	-	-	-	.54	-
.63 - .37	-	-	-	-	-	-	-	-	-	-	-	.64	-	-	-	.43
.62 - .38	-	-	-	-	1.00	-	-	-	.56	-	-	-	.50	.54	-	-
.61 - .39	-	-	-	-	-	.71	-	-	-	.60	.64	-	-	-	.43	-
.60 - .40	-	1.00	-	-	-	-	.75	-	-	-	-	.50	.54	-	-	.47
.59 - .41	-	-	-	-	-	-	-	.56	-	-	-	-	-	.43	-	-
.58 - .42	-	-	1.00	-	-	-	-	-	.60	.64	-	-	-	-	.47	.38

Table 3 (cont.)

Difficulty	Total Group Size															
	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38
.57 - .43	-	-	-	1.00	.71	-	-	-	-	-	.50	.54	-	-	-	-
.56 - .44	-	-	-	-	-	.75	-	-	-	-	-	-	.43	.47	.38	-
.55 - .45	-	-	-	-	-	-	.56	-	-	-	-	-	-	-	-	.41
.54 - .46	-	-	-	-	-	-	-	.60	.46	.50	-	-	-	-	-	-
.53 - .47	-	-	-	-	-	-	-	-	-	-	.54	.43	.47	.38	.41	.44
.52 - .48	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.51 - .49	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.50 - .50	1.00	1.00	.67	.71	.75	.56	.60	.46	.50	.54	.43	.47	.38	.41	.33	.37

Table 3 (cont.)

Difficulty	Total Group Size															
	30	42	44	46	48	50	52	54	56	58	60	62	64	66	68	70
.99 - .01	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.98 - .02	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.97 - .03	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.94 - .04	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.95 - .05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.94 - .06	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.93 - .07	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.00	1.00
.92 - .08	-	-	-	-	-	-	-	-	-	-	1.00	1.00	1.00	1.00	-	-
.91 - .09	-	-	-	-	-	-	-	1.00	1.00	1.00	-	-	1.00	1.00	1.00	1.00
.90 - .10	-	-	-	-	1.00	1.00	1.00	-	-	1.00	1.00	1.00	1.00	-	1.00	1.00
.89 - .11	-	1.00	1.00	1.00	-	-	-	1.00	1.00	-	-	1.00	-	1.00	-	.75
.88 - .12	-	-	-	-	-	1.00	1.00	-	-	1.00	1.00	-	-	.75	.75	-
.67 - .13	1.00	-	-	1.00	1.00	-	-	1.00	1.00	-	.75	.75	.75	-	.78	.78
.86 - .14	-	1.00	1.00	-	-	.71	.71	-	.74	.74	-	-	.78	.78	-	.60
.85 - .15	1.00	-	-	.71	.71	-	.75	.75	-	-	.78	.78	-	.60	.60	-
.84 - .16	-	-	.71	-	-	.75	-	-	.78	.78	-	.60	.60	-	.64	.64
.83 - .17	-	.71	-	.75	.75	-	.78	.73	-	.60	.60	-	.64	.64	-	.67
.82 - .18	.71	-	.75	-	-	.78	-	-	.60	-	.64	.64	-	.67	.67	-
.81 - .19	-	.75	-	-	.78	-	.60	.60	-	.64	-	.67	.67	-	.54	.54
.80 - .20	.75	-	.78	.78	-	.60	-	.64	.64	-	.67	-	.54	.53	-	.57
.79 - .21	-	.78	-	-	.60	-	.64	-	.67	.67	-	.54	-	.57	.57	.47

Table 3 (cont.)

Difficulty	Total Group Size															
	40	42	44	46	48	50	52	54	56	58	60	62	64	66	68	70
.78 - .22	-	-	.70	.60	-	.64	-	.50	-	.54	.54	-	.57	-	.47	-
.77 - .23	.78	-	-	-	.64	-	.50	-	.54	-	.57	.57	.47	.47	-	.50
.76 - .24	-	.60	-	.64	-	.50	-	.54	-	.57	-	.47	-	.50	.50	.41
.75 - .25	.60	-	.64	-	.60	-	.54	-	.57	-	.47	.50	.50	-	.41	-
.74 - .26	-	.64	-	.50	-	.54	-	.57	-	.47	-	-	-	.41	-	.44
.73 - .27	-	-	.50	-	.54	-	.57	-	.47	-	.50	.41	.41	.44	.44	.47
.72 - .28	.64	-	-	.54	-	.57	-	.48	-	.50	.41	-	.44	-	.47	-
.71 - .29	-	.50	-	-	.57	-	.47	-	.50	.41	-	.44	-	.47	.40	.40
.70 - .30	.50	-	.54	.57	-	.47	-	.50	.41	-	.44	-	.37	.40	-	.43
.69 - .31	-	.54	-	-	.47	-	.50	-	-	.44	-	.37	.40	-	.43	.36
.68 - .32	-	-	.57	-	-	.50	-	.41	.44	-	.37	.40	-	.43	.35	-
.67 - .33	.54	.57	-	.47	.50	-	.41	.44	-	.37	.40	-	.43	.36	-	.39
.66 - .34	-	-	.47	-	-	.41	-	-	.37	-	-	.43	.36	-	.39	.33
.65 - .35	.43	-	-	.50	.41	-	.44	.37	-	.40	.43	-	-	.39	.33	-
.64 - .36	-	.47	.50	-	-	.44	-	-	.40	.43	-	.36	.39	.33	-	.36
.63 - .37	-	-	-	.42	-	-	.37	.40	-	-	.36	.39	-	-	.36	.31
.62 - .38	.47	.50	-	-	.44	.37	-	-	.33	.36	.39	-	.33	.36	.31	-
.61 - .39	-	-	.41	.44	-	-	.40	.33	-	-	-	.33	.36	.31	-	.33
.60 - .40	.50	-	-	-	.37	.40	.33	-	.36	.39	.33	.36	-	-	.33	.29
.59 - .41	-	.41	.44	.37	-	-	-	.36	-	.33	-	-	.31	.33	.29	.31
.58 - .42	-	-	-	-	.40	.33	.36	-	.39	-	.36	.31	.33	.29	-	-

Table 3 (cont.)

Difficulty	Total Group Size															
	40	42	44	46	48	50	52	54	56	58	60	62	64	66	68	70
.57 - .43	.41	.44	.37	-	-	-	-	.30	-	.36	.31	-	-	-	.31	.27
.56 - .44	-	-	-	.40	.33	.36	.30	.33	.33	-	-	.33	.29	.31	.27	.29
.55 - .45	.44	.37	-	-	-	-	-	-	-	.31	.33	.29	.31	-	-	-
.54 - .46	-	-	.40	.33	.36	.30	.33	.36	.36	-	-	-	-	.27	.30	.25
.53 - .47	-	-	-	-	-	-	-	-	.31	.33	.29	.31	.27	.29	.25	.27
.52 - .48	.37	.48	.33	.36	.30	.33	.28	.31	-	.29	.31	.27	.29	-	-	-
.51 - .49	-	-	-	-	-	-	-	-	.33	-	-	-	-	.25	.27	.24
.50 - .50	.40	.33	.37	.30	.33	.28	.31	.33	.29	.31	.27	.29	.25	.27	.24	.26

Table 3 (cont.)

Difficulty	Total Group Size														
	72	74	76	78	80	82	84	86	88	90	92	94	96	98	100
.99 - .01	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.98 - .02	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.97 - .03	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.96 - .04	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
.95 - .05	-	-	-	-	-	-	-	-	-	-	1.00	1.00	1.00	1.00	-
.94 - .06	-	-	-	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00
.93 - .07	1.00	1.00	1.00	-	-	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
.92 - .08	1.00	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	1.00	-	.75	.75	1.00
.91 - .09	-	-	1.00	1.00	1.00	1.00	-	.75	.75	.75	.75	.75	.78	.78	.75
.90 - .10	1.00	1.00	-	.75	.75	.75	.75	-	.78	.78	.78	.78	.60	.60	.78
.89 - .11	.75	.75	.75	-	.78	.78	.78	.78	.60	.60	.60	.60	-	.64	.60
.88 - .12	-	.78	.78	.78	-	.60	.60	.60	-	.64	.64	.64	.64	.67	.64
.87 - .13	.78	-	.60	.60	.60	.64	.64	.64	.64	.67	.67	.67	.67	.54	.67
.86 - .14	.60	.60	-	.64	.64	-	.67	.67	.67	.54	.54	.54	.54	.50	.54
.85 - .15	.64	.64	.64	.67	.67	.67	-	.54	.54	-	.57	.57	.57	.47	.57
.84 - .16	-	.67	.67	-	.54	.54	.54	.57	.57	.57	.47	.47	.47	.50	.47
.83 - .17	.67	-	.54	.54	-	.57	.57	.47	.47	.47	.50	.50	.50	.53	.50
.82 - .18	.54	.54	.57	.57	.57	.47	.47	-	.50	.50	-	.53	.52	.44	.53
.81 - .19	.57	.57	-	.47	.47	-	.50	.50	.53	.53	.53	.44	.44	.47	.44
.80 - .20	-	.47	.47	-	.50	.50	.53	.53	-	.44	.44	.47	.47	.40	.47
.79 - .21	.47	-	.50	.50	.41	.53	.44	.44	.44	.47	.47	.40	.40	.43	.40

Table 3 (cont.)

	Total Group Size														
Difficulty	72	74	76	78	80	82	84	86	88	90	92	94	96	98	100
.78 - .22	.50	.50	.41	.41	-	.44	-	.47	.47	.40	.40	.43	.43	.36	.43
.77 - .23	-	.41	-	.44	.44	.47	.47	.40	.40	.43	.43	.35	.36	-	.36
.76 - .24	.41	.44	.44	.47	.47	.40	.40	.43	.43	.35	.36	-	.39	.39	.39
.75 - .25	.44	-	.47	-	.40	-	.43	-	.35	-	.39	.39	.33	.33	.42
.74 - .26	.47	.47	.40	.40	.43	.43	.36	.36	.36	.39	.33	.33	.36	.39	.39
.73 - .27	-	.40	-	.43	-	.36	.39	.39	.33	.33	.36	.36	.39	.39	.33
.72 - .28	.40	.43	.43	.36	.36	.39	-	.33	.36	.36	.39	.39	.33	.33	.36
.71 - .29	.43	-	.36	-	.39	.33	.33	.36	-	.39	.33	.33	.36	.36	.31
.70 - .30	.36	.36	.39	.39	.33	-	.36	.31	.39	.33	.36	.26	.31	.31	.33
.69 - .31	-	.39	-	.33	.36	.36	.31	.33	.33	.36	-	.31	.33	.33	.29
.68 - .32	.39	.33	.33	.36	-	.31	.33	-	.36	.31	.31	.33	.29	.29	.31
.67 - .33	.33	-	.36	.31	.31	.33	.36	.36	.31	.33	.33	.29	.31	.31	.27
.66 - .34	-	.36	.31	-	.33	.36	-	.31	.33	.29	.29	.31	.27	.27	.29
.65 - .35	.36	.31	-	.33	.36	.31	.31	.33	.29	-	.31	.27	.29	.29	.26
.64 - .36	.31	-	.33	.29	.31	-	.33	.29	.31	.31	.27	.29	.26	.26	.27
.63 - .37	-	.33	.29	.31	-	.33	.29	.31	-	.27	.29	.26	-	.28	.24
.62 - .38	.33	.29	.31	-	.33	.29	.31	.27	.27	.29	.26	.28	.28	.24	.26
.61 - .39	.29	.31	-	.33	.29	.31	.27	-	.29	.26	.28	.24	.24	.26	.28
.60 - .40	.31	-	.29	.29	.31	.27	-	.29	.26	.28	.24	.26	.26	.23	.25
.59 - .41	-	.27	.29	.31	.27	-	.29	.26	.23	.24	.26	-	.23	.25	.27
.58 - .42	.27	.29	.31	.27	-	.29	.26	.28	.24	.26	.23	.23	.25	.27	.24

Table 3 (cont.)

Difficulty	Total Group Size														
	72	74	76	78	80	82	84	86	88	90	92	94	96	98	100
.57 - .43	.29	.25	.27	-	.29	.26	.28	.24	.26	.23	-	.25	.22	.24	.26
.56 - .44	.25	-	-	.29	.26	.28	.24	.26	.23	.25	.25	.22	.24	.26	.23
.55 - .45	-	.27	.29	.26	.28	.24	.26	.23	-	-	.22	.24	.26	.23	.24
.54 - .46	.27	.29	.26	.28	.24	.26	.23	-	.25	.22	.24	.21	.23	.24	.22
.53 - .47	.24	.26	.28	.24	-	-	-	.25	.22	.24	.21	.23	.24	.22	.23
.52 - .48	-	-	-	-	.26	.23	.25	.22	.24	.21	.23	.20	.22	.23	.21
.51 - .49	.26	.28	.24	.26	.23	.25	.22	.24	.21	.23	.20	.22	.23	.21	.22
.50 - .50	.22	.24	.26	.23	.25	.22	.24	.21	.23	.20	.22	.19	.21	.22	.20

Total Test Efficiency

This paper has been concerned primarily with the computation of the e-index at an item level. This does not, however, preclude its use as a total test statistic. Just as one can talk of total test difficulty so also can one talk of total test efficiency. In this section the equations for computing total test efficiency will be discussed. No attempt will be made to interpret the total test efficiency index other than the cursory definition that follows from it computationally.

Assume some test composed of j items. Then N_j individuals will respond to item j. The number of individuals in either g_1 or g_2 will be denoted as n_j for the j^{th} item and ($2n_j = N_j$). The difficulty and discrimination indices for the j^{th} item may be denoted as a_j and d_j respectively. The total number of possible discriminations that may be made by the j^{th} item is $\frac{N_j^2}{4}$. The absolute maximum frequency of proper discriminations that is possible for the j^{th} item, b_j^*, is given as:

$$\{b_j^*\} = N_j a_j;$$

$$\{b_j^*\} = 2n_j a_j. \qquad [34]$$

The frequency of proper discriminations less improper discriminations, $\{b\}$, for the j^{th} item is given as:

$$\{b_j\} = n_j b_j. \qquad [35]$$

Inasmuch as efficiency is defined, at an item level, as the ratio of observed proper discriminations less improper discriminations to the maximum possible number of proper discriminations, for a given difficulty index and group size, assume a similar definition for total test efficiency. Let total test efficiency be represented by the ratio of total observed proper discriminations less improper discriminations, $\{B\}$, to the maximum possible number of proper discriminations for the total test, $\{\underline{B}^*\}$. Let $\underline{E}$ represent total test efficiency, assuming that the difficulty of all items is less than .50

$$\underline{E}_1 = \frac{\{\underline{B}\}}{\{\underline{B}^*\}} \qquad [36]$$

where

$$\underline{B} = \sum_{i=1}^{j} \underline{n}_i \underline{b}_i \qquad [37]$$

and

$$\{\underline{b}^*\} = 2 \sum_{i=1}^{j} \underline{n}_i \underline{a}_i ; \qquad [38]$$

thereby allowing equation 36 to be rewritten in computational form as,

$$\underline{E}_1 = \frac{\sum_{i=1}^{j} \underline{n}_i \underline{b}_i}{2 \sum_{i=1}^{j} \underline{n}_i \underline{a}_i} . \qquad [39]$$

Such an index is indicative of the proportion of "quality" discriminations made by a total test given that all items have a difficulty less than .50. For items having a difficulty index greater than .50 equation 38 is modified to define

$$\{\underline{B}^*\} = 2 \sum_{i=1}^{j} \underline{n}_i (1-\underline{a}_i) \qquad [40]$$

and equation 39 is modified to determine $\underline{E}_2$ as

$$\underline{E}_2 = \frac{\sum_{i=1}^{j} \underline{n}_i \underline{b}_i}{2 \sum_{i=1}^{j} \underline{n}_i (1-\underline{a}_i)} . \qquad [41]$$

If the $\underline{j}$ items form a mastery test then either equation 39 or 41 could be used to compute total test efficiency. If the $\underline{j}$ items form an achievement test assume that $\underline{s}$ of the items have difficulty indices less than .50 and assume that $\underline{k}$ of the items have difficulty indices greater than or equal to .50, then if the items are grouped dichotomously according to those items having difficulty indices less than .50 and those equal to or greater than .50 the general efficiency index, $\underline{E}$, of the achievement test is given by:

$$\underline{E} = \frac{\sum_{i=1}^{j} \underline{n}_i \underline{b}_i}{2\sum_{i=1}^{s} \underline{n}_i \underline{a}_i + \sum_{i=1}^{k} \underline{n}_i (1-\underline{a})} . \qquad [42]$$

Experience with this index is still growing and therefore its application to total tests is only of theoretical interest at this time. However, the systematic study of $\underline{E}$, $\underline{E}_1$ and $\underline{E}_2$ might be informative. To wit, either $\underline{E}_1$ or $\underline{E}_2$ might serve as indicators of discriminatory of difficulty homogeneity or perhaps as some sort of index of internal consistency for a mastery test. The general index, $\underline{E}$, might serve as some index of internal consistency for one achievement test.

A Pragmatic Scheme for the Use of e

This section is written explicitly for my colleagues who say that they are nonstatisticians and want something that will immediately benefit them in the interpretation of their item analyses.

Allowing a little freedom with subjectivity it is possible to set up criteria for determining easy, moderate and hard items as well as nonefficient, efficient and ideally efficient items. Let the following inequalities, based upon efficiency, $\underline{e}$, and difficulty, $\underline{a}$, serve as operational definitions of the above terms.

$0.00 \le \underline{e} \le 0.50$ nonefficient item

$0.50 \le \underline{e} \le 0.80$ efficient item

$0.80 \le \underline{e} \le 1.00$ ideally efficient item

$0.75 \le \underline{a} \le 1.00$ easy item

$0.25 \le \underline{a} \le 0.75$ moderate item

$0.00 \le \underline{a} \le 0.25$ hard item

Because exact intervals have not been made these definitions are, theoretically, not mutually exclusive, but for practical purposes they may be thought of as mutually exclusive and exhaustive with respect to difficulty and efficiency. From these definitions it is possible to categorize all items of an instrument according to Table 4.

Table 4

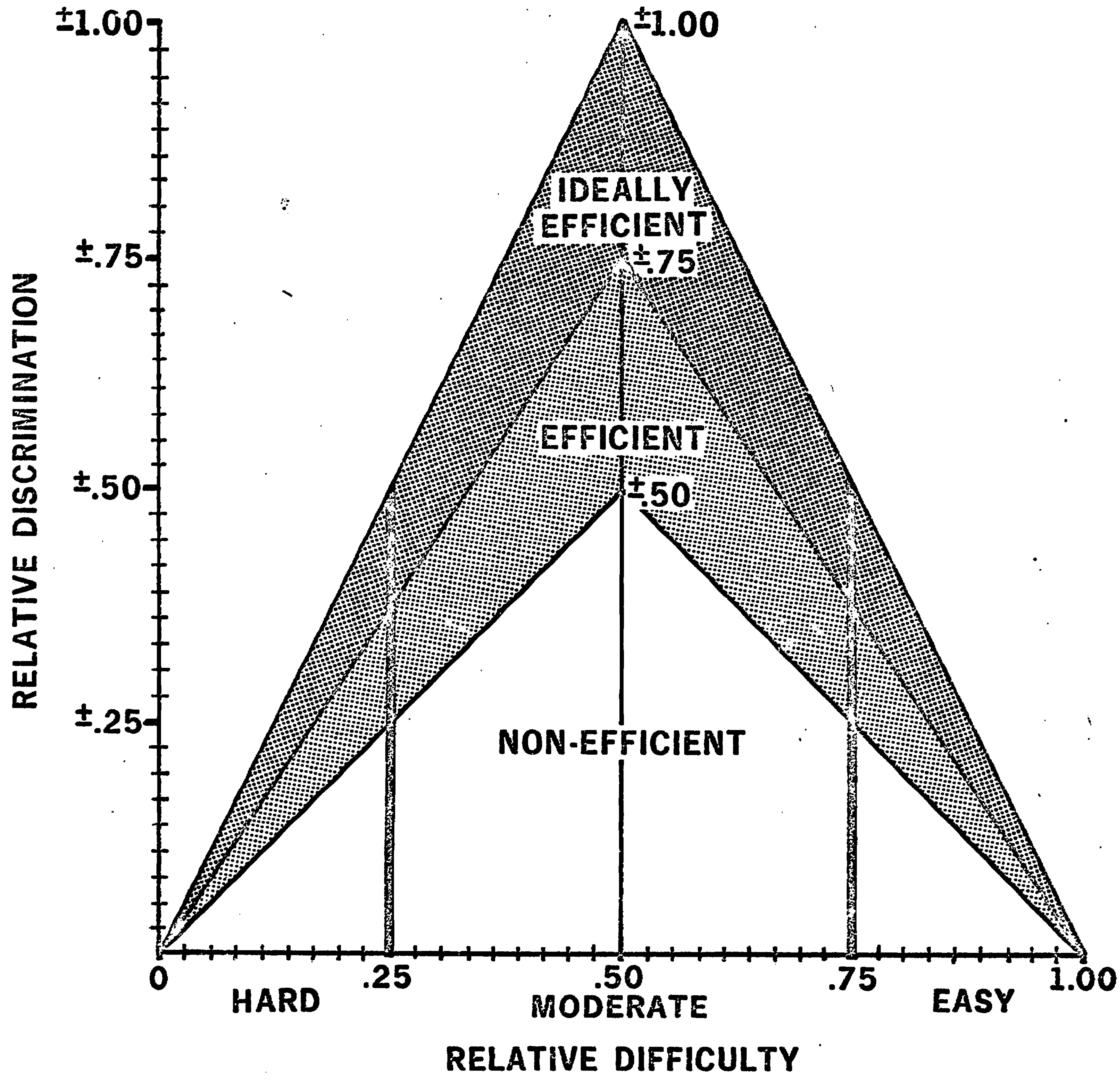
Efficiency by Difficulty Contingency Categorization of Subjective Item Types

Difficulty	Type Criterion	Efficiency Levels		
		Non-Efficient	Efficient	Ideally Efficient
Easy	Efficiency Difficulty	$0.00 \le \underline{e} \le 0.50$ $0.75 \le \underline{a} \le 1.00$	$0.50 \le \underline{e} \le 0.80$ $0.75 \le \underline{a} \le 1.00$	$0.80 \le \underline{e} \le 1.00$ $0.75 \le \underline{a} \le 1.00$
Moderate	Efficiency Difficulty	$0.00 \le \underline{e} \le 0.50$ $0.25 \le \underline{a} \le 0.75$	$0.50 \le \underline{a} \le 0.80$ $0.25 \le \underline{a} \le 0.35$	$0.80 \le \underline{e} \le 1.00$ $0.25 \le \underline{a} \le 0.75$
Hard	Efficiency Difficulty	$0.00 \le \underline{e} \le 0.50$ $0.00 \le \underline{a} \le 0.25$	$0.50 \le \underline{e} \le 0.80$ $0.00 \le \underline{a} \le 0.25$	$0.80 \le \underline{e} \le 1.00$ $0.00 \le \underline{a} \le 0.25$

Based upon Table 4 it is possible to construct within the framework of Cartesian coordinates a chart for determining item quality. In Figure 3 such a chart has been constructed. Given the difficulty and discrimination of an

Figure 3

Iconic Representation of Contingency Categorization of Subjective Item Types



item as coordinates it is possible to locate the associated item point. If the point falls within the working area then the item is a working item. Additionally it is possible to determine the relative difficulty of the working item. Although this chart was constructed as a function of the e-index, its use actually precludes the computation of such an index.

Assuming Figure 3 to be a triangular probability distribution it is possible to briefly discuss the chance probability of obtaining different item types. Note in Figure 3 that there are nine different item types. The probability of any item type's occurring by chance may be determined as the ratio of the surface area associated with each item type to the total surface area of the probability distribution. The probabilities for the nine subjective item types are reported in Table 5.

Table 5

Chance Probabilities of Subjective Item Types

Difficulty Levels	Efficiency Levels			
	Non-Efficient	Efficient	Ideally Efficient	Total
Easy	.0625	.0375	.0250	.1250
Moderate	.3750	.2250	.1500	.7500
Hard	.0625	.0375	.0250	.1250
Total	.500	.300	.200	1.0000

As previously noted the concept of item types is subjective as are the arbitrary numerical criteria for defining them. It is possible to construct a more detailed chart than Figure 3, as well as a more detailed table of chance probabilities to go with it. However, the objective of this small section was one of establishing a pragmatic schema for the use of e, which was done.

Conclusion and Summary

The conclusion to this paper will provide an efficiency analysis of an item given a difficulty index of .30, a discrimination index of .29 and a sample size of 14.

(a) By equations 5 and 13 the general efficiency index of the items is approximately .42.

(b) Given from above that ($\underline{e}$ = .42) then the probability of an individual being a member of group one given that the individual made a positive response is (.71), from either equation 28 or Table 2.

(c) The odds are 2.45 to 1.0 that a person responding positively to the item will be a member of group one, from Table 2.

(d) The value of $\underline{e}_2$ may be computed as .18 utilizing equation 32-a.

(e) Given that ($\underline{e}_2$ = .18) then the probability of an individual being a member of group two given that the individual made a negative response is .59, from either equation 31 or Table 2.

(f) The odds are 1.439 to 1.0 that a person responding negatively to the item will be a member of group two, from Table 2.

(g) According to Table 2 the observed $\underline{e}$ is not significant at the .05 level.

(h) According to Table 4 or Figure 3 this item is a "moderate, non-efficient item."

(i) From Table 5 it may be inferred that the probability of getting a moderate non-efficient item, by chance, is .375, more probable than any other type.

The nine interpretations given above are with respect to a single item. Had more items been included in the analysis it would have been possible to make comparative statements as well as relative judgements of the items.

This paper presents the basic aspects of the efficiency index. Whether or not the index will replace or supplement the traditional discrimination index remains to be seen. However, it would seem that the general efficiency index has many more statistically compelling properties than the traditional discrimination index.

References

Cureton, E. E. The upper and lower twenty-seven per-cent rule. *Psychometrika*, 1957, 22(3), 293-296.

Fisher, R. A. The logic of inductive inference. *Journal of Royal Statistical Society*, 1935, 98(1), 39-54.

Henrysson, S. Gathering, analyzing, and using data on test items. In R. L. Thorndike (Ed.), *Educational measurement*. (2nd ed.) Washington, D. C.: American Council on Education, 1971. Pp. 30-59.

Hofmann, R. The Concept of Efficiency in Item Analysis (in preparation), 1972.

Johnson, A. P. Notes on a suggested index of time validity: the U-L index. *Journal of Educational Psychology*, 1951, 42, 499-504.

Kelley, T. L. The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*. 1939, 40, 17-19.

Pearson, K. Experimental discussion of the (χ^2, P) test for goodness of fit. *Biometrika*, 1932, 24, 351-381.